A Semi-supervised Generalized VAE (ss-gVAE) Framework for Abnormality Detection using One-Class Classification

Renuka Sharma[1,2], Satvik Mashkaria[1], and Suyash P. Awate[1]

[1]Computer Science and Engineering Department, Indian Institute of Technology (IIT) Bombay, Mumbai. [2]IITB-Monash Research Academy, Mumbai.

Motivation and Introduction

- We propose a novel semi-supervised variational learning DNN framework leveraging generalized-Gaussian (GG) models on both
  - encodings in the latent space and
  - the reconstructed image.
- We propose a generalized version of the variational autoencoder (VAE) framework (gVAE) leading to data-adaptive modeling subsuming
  - robust modeling through the GG's shape parameters and
  - uncertainty-aware modeling through the GG's scale parameters.
- We further extend unsupervised gVAE to a semi-supervised framework ss-gVAE that can leverage some outlier data during training to improve performance.
- To enable backpropagation-based optimization, we propose a reparameterization of the GG.
- We show improvements on the real-world datasets along with a synthetic image set.

Proposed: Generalized Variational Autoencoder (gVAE)

gVAE is our proposed solution for unsupervised anomaly detection where we have training data from the inlier class (X) only. We minimize the loss for the normal data points using the encoder-decoder based network using GG modeling and variational learning: $\arg\min_{\theta^{\mathcal{ED}}} \mathcal{L}_{\text{normal}}(X;\theta^{\mathcal{ED}})$, where $\mathcal{L}_{\text{normal}}(X;\theta^{\mathcal{ED}})$ depicts the loss function for the normal data points.

$$\mathcal{L}_{\text{normal}}(X;\theta^{\mathcal{ED}}) := \frac{1}{NI}\sum_{n=1}^{N}\sum_{i=1}^{I}\Big[0.5\|z_{ni}(X_n,\theta^{\mathcal{E}})\|_2^2 + \log \mathcal{G}(z_{ni}(X_n;\theta^{\mathcal{E}});\mathcal{E}^{\mu}(X_n;\theta^{\mathcal{E}}),\mathcal{E}^{\alpha}(X_n;\theta^{\mathcal{E}}),\mathcal{E}^{\beta}(X_n;\theta^{\mathcal{E}})) - \log \mathcal{G}(X_n;\mathcal{D}^{\mu}(Z_n;\theta^{\mathcal{D}}),\mathcal{D}^{\alpha}(Z_n;\theta^{\mathcal{D}}),\mathcal{D}^{\beta}(Z_n;\theta^{\mathcal{D}}))\Big],$$

where X is the input image or random field, $\theta^{\mathcal{ED}}$ are the parameters for encoder-decoder, N is the number of normal data points, $\mathcal{G}$ represents GG, and z_{ni} is the i-th independent draw from the conditional PDF of the latent-space encoding Z. The second term in the expression works on the latent-space encoding and the third on reconstructed image. (Please refer to the paper for more details.)

Proposed: Semi-supervised gVAE (ss-gVAE)

ss-gVAE is an extension of gVAE which takes in a fraction of outliers Y along with the inliers X as part of the training set which helps in refining the normal data points' boundary. We have separate encoder-decoder based losses for the normal and abnormal data points and they are weighted by the hyperparameter η and we minimize the loss as follows: $\arg\min_{\theta^{\mathcal{ED}}} \eta\mathcal{L}_{\text{normal}}(X;\theta^{\mathcal{ED}}) - (1-\eta)\mathcal{L}_{\text{abnormal}}(Y;\theta^{\mathcal{ED}})$, $\mathcal{L}_{\text{abnormal}}$ is the loss function for the abnormal data points, similar to $\mathcal{L}_{\text{normal}}$ except that it works on Y instead of X. (Please refer to the paper for more details.)

Thanks to the Infrastructure Facility for Advanced Research and Education in Diagnostics grant funded by Department of Biotechnology, Government of India (BT/INF/22/SP23026/2017).

Proposed Learning Framework

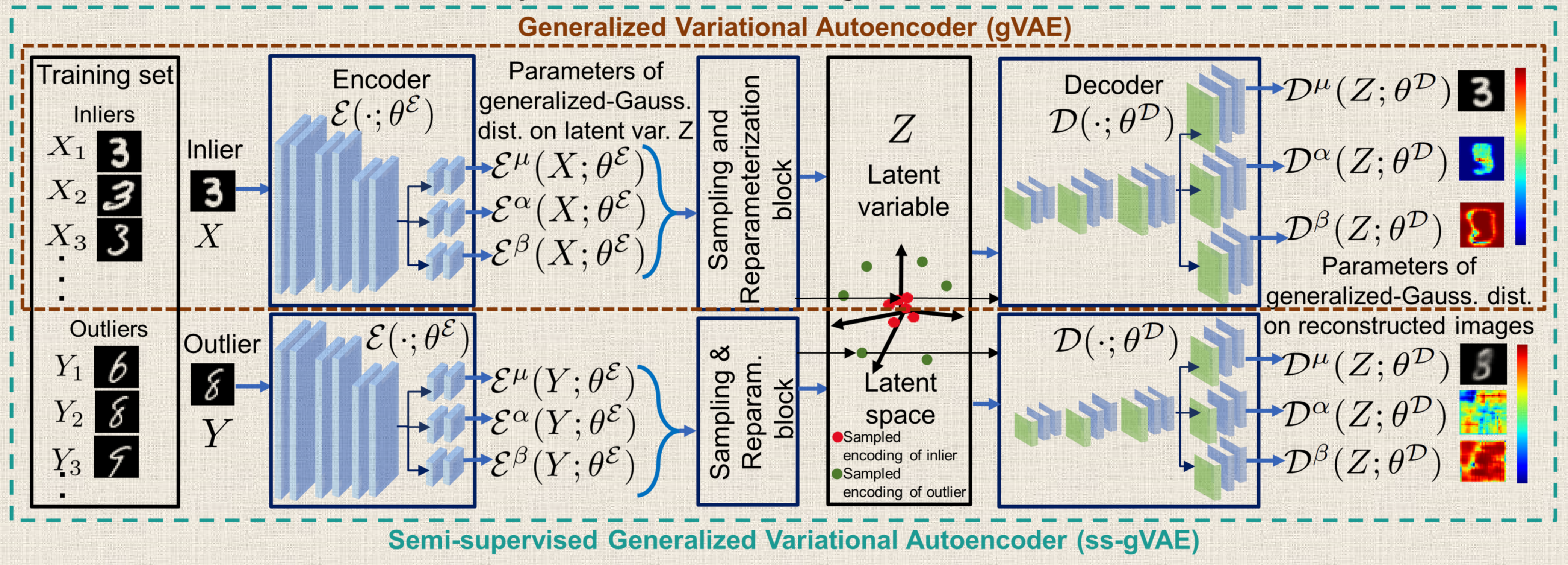


Fig: Framework for ss-gVAE extending gAVE. In this example using MNIST data, we treat images of digit 3 as inliers and images of all other digits as outliers. Given a test image U, we classify it as inlier/outlier based on the norm of the encoder-mapped mean vector.

Datasets used

- Toy datasets (MNIST, Fashion-MNIST, and CIFAR10)
- 10 datasets from MVTec Anomaly Detection (consisting of texture and object categories, each has sub-categories of anomalies), taking patches of size 64x64
- Dataset of Malaria Infected Cells comprising of red blood cells (RBCs) images that are labelled as non-infected (normal) and infected, taking patches of size 170x170 as 170 is the average cell diameter
- Synthetic dataset the intensities in each color channel are drawn independently from Gaussian distributions with the mean and the scale parameters chosen randomly within specified ranges for the normal and abnormal classes.

Qualitative results

Quantitative results

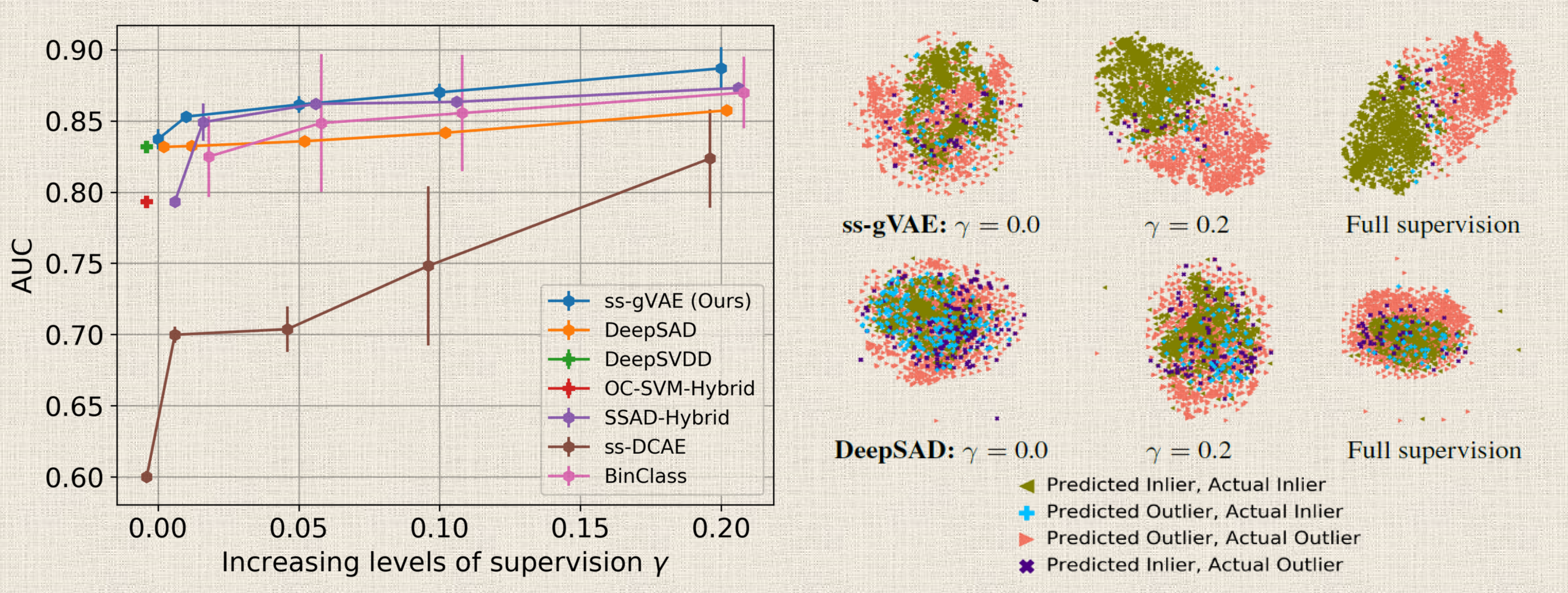
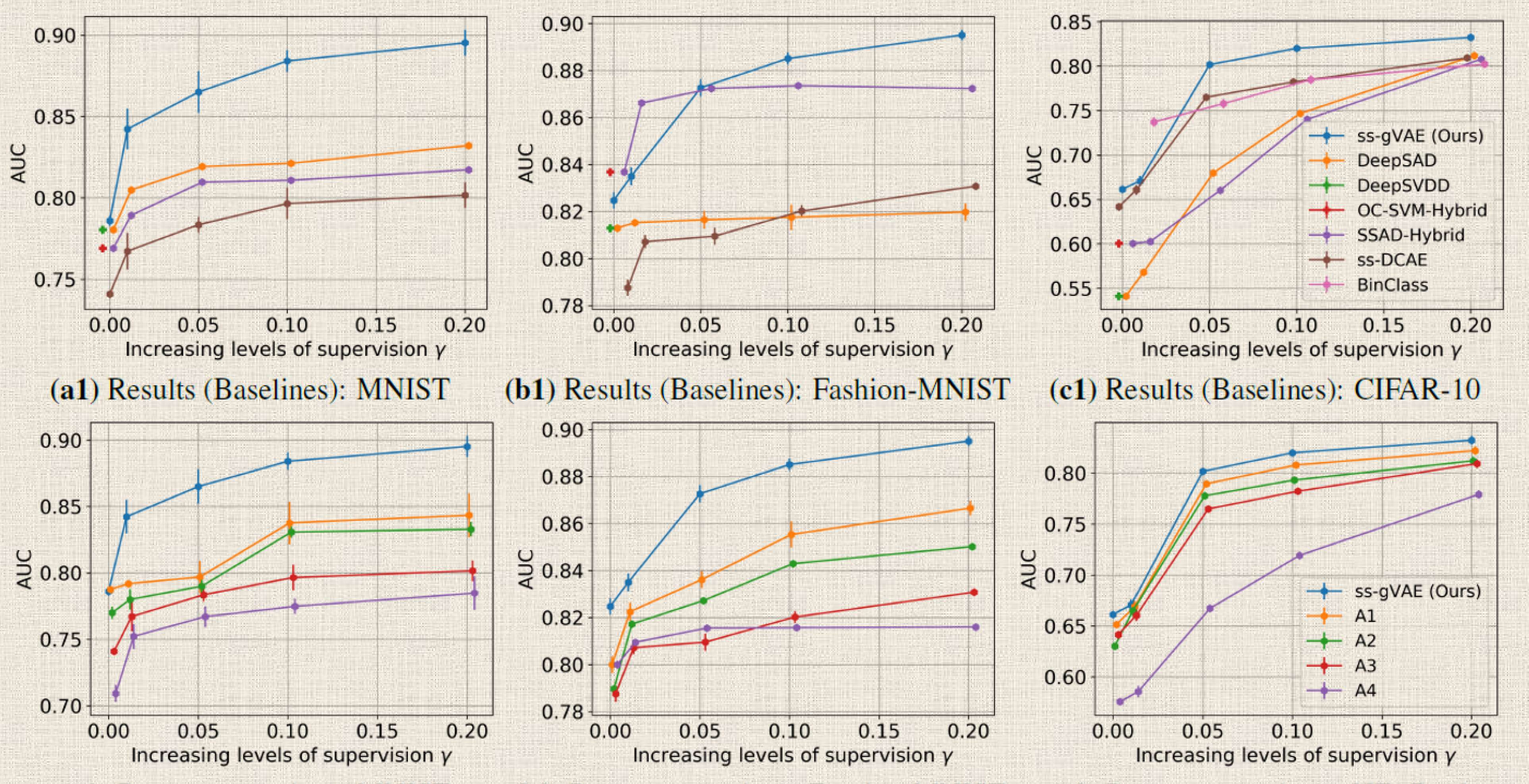


Fig: Results on Malaria dataset

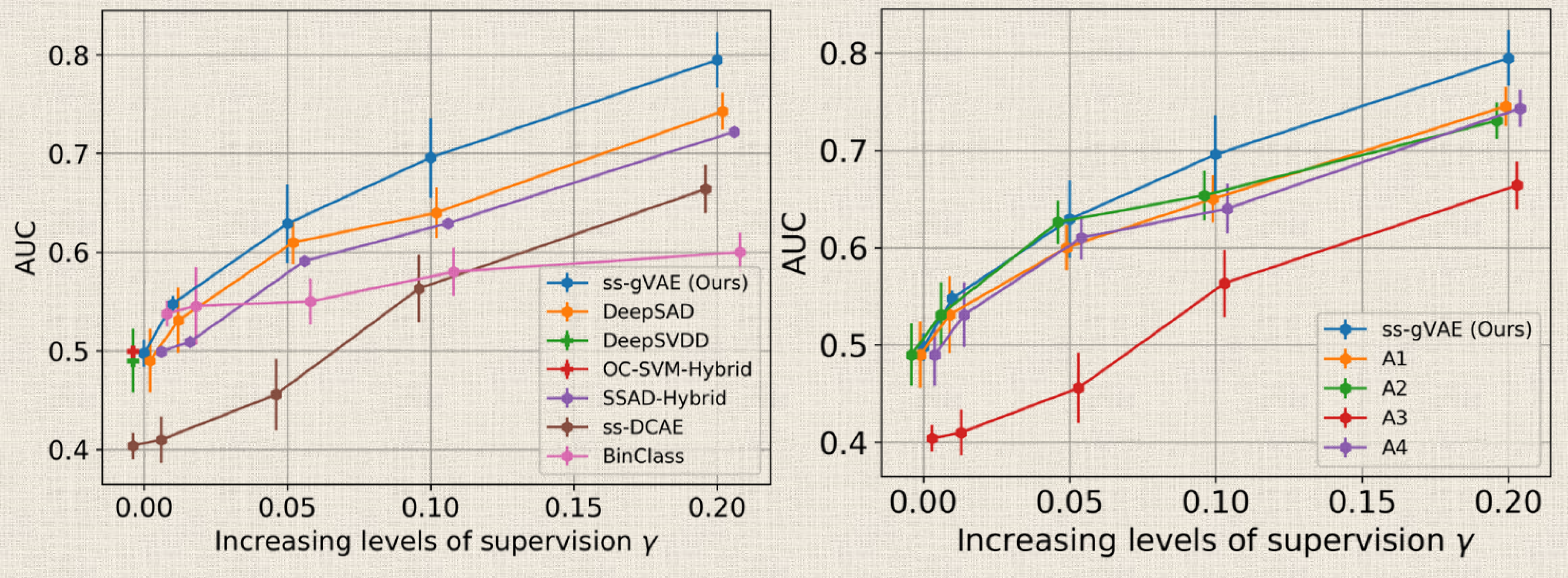


(a1) Results (Baselines): MNIST (b1) Results (Baselines): Fashion-MNIST (c1) Results (Baselines): CIFAR-10

(a2) Results (Ablation): MNIST (b2) Results (Ablation): Fashion-MNIST (c2) Results (Ablation): CIFAR-10

Fig: Results on toy dataset (Row1: comparison with the baselines, Row2: ablation studies). In all the datasets, we keep one of the classes as normal and the rest abnormal in order to simulate one-class classification.

Fig: Sample images from MVTec dataset's carpet category. A patch is labelled anomalous if at least 50% of the corresponding ground truth mask contains the anomaly. There are 5 kinds of anomalous classes in the Carpet category, half of which we keep as novel test anomalies.



(a) MVTec: Comparison with Baselines (b) MVTec: Ablation Studies

Fig: Results on MVTec dataset's Carpet category